FILTERING IN HYBRID DYNAMIC BAYESIAN NETWORKS

Morten Nonboe Andersen, Rasmus Ørum Andersen[*]

Technical University of Denmark
Informatics and Mathematical Modelling (IMM)
DK-2800 Lyngby, Denmark

Kevin Wheeler

NASA Ames Research Center
Computational Sciences Division
Moffett Field, CA 94035-1000, USA

ABSTRACT

We demonstrate experimentally that inference in a complex hybrid Dynamic Bayesian Network (DBN) is possible using the 2-Time Slice DBN (2T-DBN) from [Koller & Lerner, 2000] to model fault detection in a watertank system. In [Koller & Lerner, 2000] a generic Particle Filter (PF) is used for inference. We extend the experiment and perform approximate inference using The Extended Kalman Filter (EKF) and the Unscented Kalman Filter (UKF). Furthermore, we combine these techniques in a 'non-strict' Rao-Blackwellisation framework and apply it to the watertank system. We show that UKF and UKF in a PF framework outperform the generic PF, EKF and EKF in a PF framework with respect to accuracy and robustness in terms of estimation RMSE. Especially we demonstrate the superiority of UKF in a PF framework when our beliefs of how data was generated are wrong. We also show that the choice of network structure is very important for the performance of the generic PF and the EKF algorithms, but not for the UKF algorithms. Furthermore, we investigate the influence of data noise in the watertank simulation. Theory and implementation is based on the theory presented in [v.d. Merwe et al., 2000].

1. INTRODUCTION

Currently, most of the problems presented in literature are limited to static Bayesian networks, networks with discrete variables or with linear relations between variables. In this paper, we assume a Markovian, stationary model and setup a 2T-DBN in which nodes given at time t are dependent only on variables at time t and $t-1$. This allows time varying relations between variables (however the network *structure* is constant over time). We keep all observations in their 'true' domain leading to a model with both discrete and continuous valued nodes. Our model also has linear as well as non-linear relations. Showing how to do inference in these very general models is very important as they allow for many real-life problems to be modelled. Hence, we need to compare the performance of different algorithms, investigate their sensitiveness to the choice of network structure and find out if the algorithms fail when we do not know the true variable relations or when we have a noisy environment. In this paper, bold face symbols indicate a vector or a matrix and standard face symbols are scalars.

2. FILTERING IN BAYESIAN NETWORKS

Converting a multivariate Gaussian distribution into a Bayesian network by ordering the variables $x_1, ..., x_n$ topologically, the distribution of a child conditioned on its parents is computed as

$$P(x_i|x_1, ..., x_{i-1}) = \mathcal{N}\left(x_i; \beta_{i,0} + \sum_{j=1}^{i-1} \beta_{i,j} x_j, \sigma_i^2\right).$$

An edge from x_j to $x_i (1 \leq j < i)$ corresponds to $\beta_{i,j} \neq 0$. In general, filtering is the problem of estimating the state of a system using a set of on-line observations. We solve this problem by modelling the evolution of the system consisting of a *state transition or state process* model $p(\boldsymbol{x}_t|\boldsymbol{x}_{t-1})$ and *a state measurement* model $p(\boldsymbol{y}_t|\boldsymbol{x}_t)$ and the noise on the measurements. $\boldsymbol{x}_t \in \Re^{n_x}$ are the states (hidden variables or parameters) of the system at time t and $\boldsymbol{y}_t \in \Re^{n_y}$ are the observations. For example, non-linear, non-Gaussian models can be expressed as $\boldsymbol{x}_t = \boldsymbol{f}(\boldsymbol{x}_{t-1}, \boldsymbol{v}_{t-1})$ and $\boldsymbol{y}_t = \boldsymbol{h}(\boldsymbol{x}_t, \boldsymbol{n}_t)$ with $\boldsymbol{v}_t \in \Re^{n_v}$ being the process noise and $\boldsymbol{n}_t \in \Re^{n_n}$ the measurement noise.

The EKF is a minimum mean-square-error (MMSE) estimator based on the Taylor series expansion of the non-linear functions $\mathbf{f}$ and $\mathbf{h}$ around the estimates $\bar{\mathbf{x}}_{t|t-1}$ of the states $\mathbf{x}_t$, e.g.

$$\boldsymbol{f}(\boldsymbol{x}_t, \boldsymbol{v}_t) = \boldsymbol{f}(\bar{\boldsymbol{x}}_{t|t-1}, \bar{\boldsymbol{v}}_{t|t-1}) + \frac{\partial \boldsymbol{f}(\boldsymbol{x}_t, \boldsymbol{v}_t)}{\partial \boldsymbol{x}_t}\bigg|_{(\boldsymbol{x}_t = \bar{\boldsymbol{x}}_{t|t-1})}$$

$$(\boldsymbol{x}_t - \bar{\boldsymbol{x}}_{t|t-1}) + \frac{\partial \boldsymbol{f}(\boldsymbol{x}_t, \boldsymbol{v}_t)}{\partial \boldsymbol{v}_t}\bigg|_{(\boldsymbol{v}_t = \bar{\boldsymbol{v}}_{t|t-1})}(\boldsymbol{v}_t - \bar{\boldsymbol{v}}_{t|t-1}) + \dots$$

The UKF [Julier & Uhlmann, 1997] is a recursive MMSE estimator that does not approximate the non-linear process and measurement models, but makes a Gaussian approximation of the distribution of the state random variable. When this variable is propagated through the true non-linear system, it captures the true mean and covariance to the second order for any non-linearity.

PF represents a generalization of Monte Carlo methods for a dynamic process. The particles are weighted recursively using importance weights $\omega_t = \omega_{t-1} \frac{p(\mathbf{y}_t|\mathbf{x}_t) p(\mathbf{x}_t|\mathbf{x}_{t-1})}{q(\mathbf{x}_t|\mathbf{x}_{0:t-1}, \mathbf{y}_{1:t})}$. In this paper our goal is to perform filtering allowing proposal (approximation of the true posterior) distributiond of the form $q(\mathbf{x}_{0:t}|\mathbf{y}_{1:t}) = q(\mathbf{x}_{0:t-1}|\mathbf{y}_{1:t-1}) q(\mathbf{x}_t|\mathbf{y}_{0:t-1}, \mathbf{y}_{1:t})$. Assuming the states follow a 1. order Markov process and that the obs. a conditionally independent given the states yields $p(\mathbf{x}_{0:t}) = p(\mathbf{x}_0) \prod_{j=1}^{t} p(\mathbf{x}_j|\mathbf{x}_{j-1})$, $p(\mathbf{y}_{1:t}|\mathbf{x}_{0:t}) = \prod_{j=1}^{t} p(\mathbf{y}_j|\mathbf{x}_j)$. In the generic PF, the transition prior $q(\mathbf{x}_t|\mathbf{x}_{0:t-1}, \mathbf{y}_{1:t}) \stackrel{\circ}{=} p(\mathbf{x}_t|\mathbf{x}_{t-1})$ is used as proposal distribution. To introduce sample variety after the selection step without affecting the validity of the approximation we may also apply a MCMC step of invariant distribution $p(\mathbf{x}_{0:t}|\mathbf{y}_{1:t})$ on each particle. We refer to this filter as $PFCM$. In this work, the EKF computes the recursive approximation of the true posterior filtering density given by $p(\mathbf{x}_t|\mathbf{y}_{1:t}) \approx p_{\mathcal{N}}(\mathbf{x}_t|\mathbf{y}_{1:t}) = \mathcal{N}\left(\bar{\mathbf{x}}_t, \hat{\mathbf{P}}_t\right)$. Using the EKF in a PF framework, a separate EKF is used to generate

and propagate a Gaussian proposal distribution for each particle $q(\mathbf{x}_t^{(i)}|\mathbf{x}_{0:t-1}^{(i)}, \mathbf{y}_{1:t}) \doteq \mathcal{N}(\mathbf{x}_t|\mathbf{y}_{1:t}), i = 1, \ldots, N$ i.e. at time $t-1$ the mean and covariance of the importance distribution for each particle are computed using the EKF equations and the new observation. This filter is known as the *Extended Kalman Particle Filter* [Doucet et al., 1998]. Using the UKF as proposal distribution generator leads to the *Unscented Filter* [v.d. Merwe et al., 2000]. These filters are abbreviated PFEKF and PFUKF resp. in this work.

3. WATERTANK SIMULATION

In this section we investigate the fault detection system presented in [Koller & Lerner, 2000]. The system is shown in Figure 1. For a more detailed presentation and more experiments using this network, please refer to [Andersen & Andersen, 2003]. In Figure 2, a
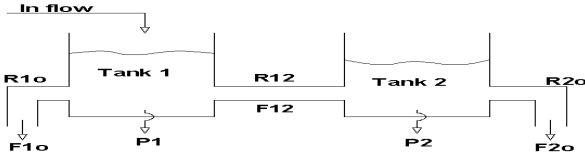

Fig. 1. Illustration of the watertank system

2T-DBN describe how flow, pressure and resistance are related in theory. Practically, the process and measurement models and the measurements themselves are noisy. Furthermore, there are three possible type of failures in the system that we would like to detect:

Measurement failure In the case of a measurement failure, the measurement becomes extremely noisy.

Pipe bursts A pipe can suddenly burst and change its resistance to some unknown value

Drifts The resistance of the pipe can drift, which gradually increases or decreases the pipes resistance
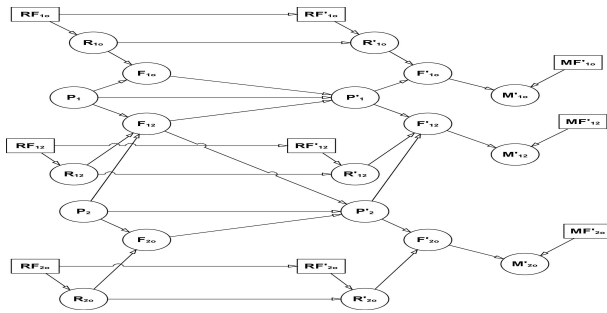

Fig. 2. DBN for the watertank system

The discrete RF nodes indicate faults in the resistance of the pipes (drifts or bursts) and the MF nodes indicate measurement failures. The P, F and R nodes are continuous valued and indicate pressure, flow and pipe resistance resp. However, as the flow is the ratio between the pressure and the resistance, we use the conductance C (reciprocal of the resistance) to avoid ratios. Finally, nodes labelled M indicate pipe flow measurements and are observable nodes. All other variables are hidden. The network has six pipe

fault variables and three measurement failure variables, leading to 32,768 different discrete states. To simplify, the pipe connecting the two tanks can not burst reducing the state space to 18,432 states. Unfortunately, this network is still far too complicated to be able to use exact inference. Sub-optimally, we would like to sample all discrete variables, which we can group into two vector-valued nodes CF_t (conductance failures) and MF_t (measurement failures), and apply exact inference on the remaining nodes, which we group into a single vector-valued node, X_t. A technique known as *Rao-Blackwellisation* (RB). The observable nodes are likewise grouped into a single vector-valued node Y_t, allowing a transformation of the fairly complicated network into a simple network as illustrated in Figure 3 using $CF_t = A \cdot CF_{t-1}$, $X_t = B(CF_t) \cdot X_{t-1}$ and $Y_t = D \cdot X_t$.

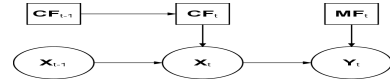**Fig. 3**. Simplified DBN for the watertank problem

Although the noise is Gaussian, the dynamics are non-linear, making it hard to integrate out X_t. Hence, we apply our approximate inference techniques EKF and UKF and call it 'non-strict' RB. To compare, we also apply a generic PF, PFMC, PFKEF and PFUKF to do inference on the continuous valued nodes,. All implementations except PF and PFMC were designed as a two-step serial process. The first process samples the discrete nodes using a generic PF algorithm, but without updating the cont. state variables. The cont. states were then estimated (for each particle) in the second process using PF, PFMC, EKF, UKF, PFEKF or PFUKF. This two-step process was used as all the EKF and UKF based algorithms were able to give good estimates of the cont. nodes based on poor estimates of the discrete nodes.

In Figure 2 we see that the flow nodes are the only hidden nodes directly connected to the observation nodes. In EKF, the Kalman gain is partly based on the Jacobian of the measurement model. Even though pressure and conductance are highly correlated with the flow, the Kalman gain thus only influences the flow estimates. Figure 4 shows the relative RMSE of the flow, conductance and pressure estimates using EKF with correct initialization of all state variables. The figure illustrates that EKF is making poor conductance and pressure estimates whereas the flow estimates are very accurate for all time steps. In the generic PF we use the transition prior as proposal distribution and all particles are
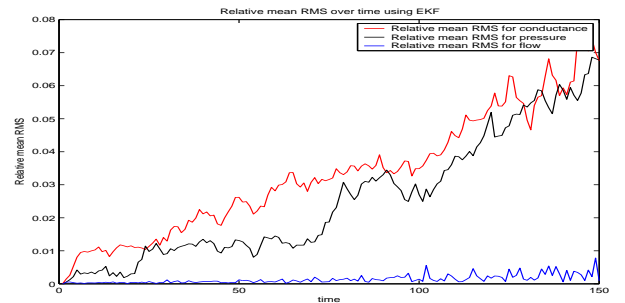

Fig. 4. Relative RMSE for conductance (red), pressure (black) and flow (blue) estimates using EKF with correct initialization.

thus weighted according to their likelihood, i.e. based on the difference between the true and predicted values of the observations. Hence, with only the flow nodes connected to the observations, a particle with accurate flow values will give a high likelihood regardless of whether the particle has poor conductance or pressure estimates. This problem is illustrated in Figure 5 which shows three different weighings of 10 particles. The actual weights used follow the weights based on the flow values and not the optimal weights. A large process noise would make this problem even worse. In UKF, the Kalman gain is based on a number of sigma
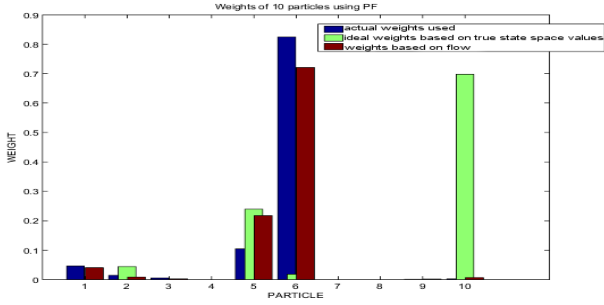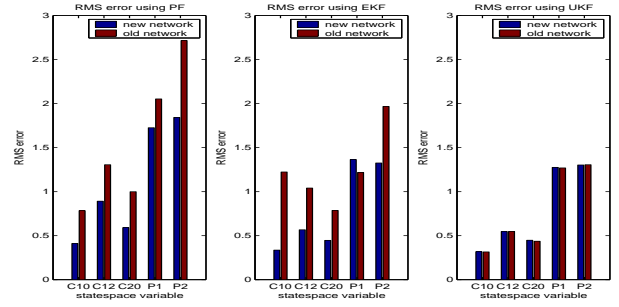

Fig. 6. Average RMSE for conductance and pressure estimates using the old network (red bars) and the new network (blue bars) using the generic PF, EKF and UKF resp.


Fig. 5. Actual weights used (blue), optimal weights based on the distance from the true cont. states (green) and weights based on the distance to the true flows (red) for 10 particles using the generic PF.

Fig. 7. Surface plots showing the RMSE (top plots) and CF/MF estimation errors (lower plots) using UKF (left plots) and PFUKF (right plots).

points that are propagated through the network using the true process and measurement models. Both pressure and conductance are highly correlated with the observation nodes, even though they are not directly connected. This property makes UKF able to update all continuous state variables. One of the objectives in the water-tank problem is to track conductance failures making accurate estimation of the conductance a crucial point. Hence, a new network was proposed by eliminating the flow nodes from the continuous state variables allowing conductance and pressure nodes to be directly connected to the observation nodes. When data is generated using the old network, noise is added to the flows making the data more noisy than data generated in the new network. To compare the two networks, the old network was used to generate data used in both networks. If the new network performs better on a data set generated by the old network, it is the obvious choice of network. Figure 6 shows the average RMSE for the conductance and pressure estimates from the two networks using PF, EKF and UKF. The results are based on 10 different data sets using 10 runs for each data set. As illustrated, the new network outperforms the old network for all continuous state mean estimates using the generic PF and EKF. In comparison, the performance of UKF does not depend on the choice of network structure.

Next, we evaluate the sensitiveness with respect to the noise by experimenting with different levels of data noise. Only UKF and PFUKF were used based on their superior performance in the previous experiments (see [Andersen & Andersen, 2003] for more experiments and details). Four different process and measurement noise levels were used, $\mathcal{N}(0, \sigma^2)$, $\sigma^2 = 0.01, 0.1, 0.2$ and 0.4 and the true noise levels were used as proposed noise levels (used in the filtering algorithms). The RMSE and the number of wrong failure estimates as a function of the noise levels are shown in Figure 7 using a time period of 100 time steps. The results were based on 20 different data sets using 10 runs for each data set. Outliers were removed. Notice the nice correlation between the RMSE and
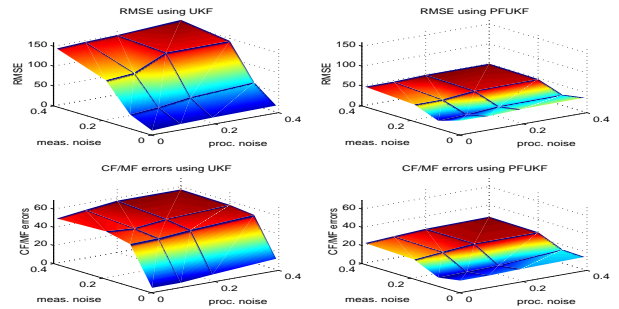
the number of wrong failure estimates for UKF (two left plots) and PFUKF (two right plots). An accurate state estimate corresponds to a small RMSE making it easier to track the failures and vice versa. Both the RMSE and the number of wrong failure estimates using UKF and PFUKF are more influenced by the level of the measurement noise than the process noise. Noisy measurements are more crucial than a noisy connection between the hidden variables as the algorithms are estimating the noisy hidden states *based on* estimates of the measurements. Furthermore, Figure 7 illustrates the relationship between UKF and PFUKF. PFUKF is doing much better than UKF for large measurement noise levels, but notice that the RMSE increases using the smallest process noise level (0.01) and is actually doing worse than UKF: When UKF makes accurate estimates, PFUKF can make matters worse by either fitting the measurement noise or sampling from a Gaussian distribution that is too wide.

In a real life simulation one of the major challenges is to come up with reasonable process and measurement models. Simulating the actual data gives us all the knowledge about the problem we need and makes us capable of providing our algorithms with optimal conditions. In this section, we change the proposed measurement model by simply adding 5% to all the flow estimates. This might affect UKF negatively leaving some space for improvement for PFUKF. PF and PFMC were also included in the experiments for the sake of comparison. 100 particles and 30 subparticles were used over a time period of 60 time steps. Table 1 shows the mean estimation RMSE and variance using 10 different data sets and 10 runs for each data set. The second column shows the average num-

ber of incorrect failure estimates. As shown, PFUKF is by far the filtering algorithm that is most capable of dealing with a wrong measurement model. PFUKF is able to move the particles towards regions of higher likelihood which reduces the RMSE and makes tracking of the discrete failure nodes easier.

Algorithm	RMS		CF/MF errors
	mean	var	
Particle Filter - generic	256	204	56.3
Particle Filter - MH move	227	187	33.8
Unscented Kalman Filter	208	43	30.1
Particle Filter - UKF proposal	178	76	23.4

Table 1. RMSE of state mean estimates (mean and var.) and average CF/MF estimation errors using a false measurement model.

In [Koller & Lerner, 2000], a generic PF is applied to the watertank problem using the network structure in Figure 2. They propose as future work a combination of a generic PF (sampling the discrete failure nodes) and a more sophisticated filter to sample the continuous variables as in this work. It is impossible to make direct comparisons between this work and the work of Koller and Lerner due to likely differences in modelling (such as pipe bursts or measurement failures). However, we have showed that the generic PF (and EKF) are highly sensitive to the choice network structure as opposed to the UKF based implementations which were superior in terms of estimation RMSE. This was further indicated using a false measurement model. Figure 8 shows the tracking of $C10$ (conductance of pipe between $Tank1$ and the outside world (only every second error bar plotted for visual reasons) using UKF together with the events that occurred during a typical simulation. Similar results were obtained for the other continuous state variables - data not shown) using UKF. We draw process and measurement noise samples from $\mathcal{N}(0, 0.5)$ and use a drifting factor of 1 (i.e. the pipe conductance changes one unit for each time step) to obtain a low SNR ratio. We present a tracking plot for UKF instead of the superior PFUKF to show that we can track the continuous variables and detect system faults very well using a very low number of particles compared to the 50000 particles used in the generic PF in [Koller & Lerner, 2000] without taking advantage of PFUKF, which is computationally more expensive than UKF.
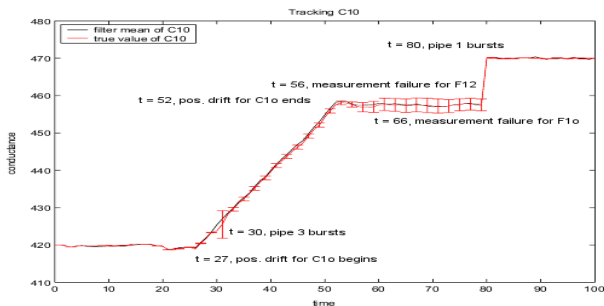

Fig. 8. Estimation conductance variable $C10$ using UKF (red line) and confidence intervals (plus, minus two standard deviations from the mean estimate) and the true conductance C10 (black line) based on ten runs with one data set using 300 particles.

4. CONCLUSION

In a 2T-DBN watertank simulation we have compared two network structures and showed that the PF and EKF estimates were network structure sensitive as opposed to the UKF estimates. Then we showed that continuous state mean estimates using UKF and PFUKF were more sensitive to changes in the measurement noise level than in the level of process noise. Large measurement noise levels made the UKF estimates poor and PFUKF was able to move the particles towards the true state, i.e. reducing the estimation RMSE. Finally, we experimented using a measurement model different from the true one by adding 5% to all the flow estimates. Again, PFUKF was capable of making more accurate state estimates than UKF and showed that PFUKF is a more reliable algorithm, when we do not know the true model relations. Furthermore, UKF and PFUKF were more accurate than PF and PFMC. Finally, we showed that we were able to track the discrete failure nodes with a fairly low number of particles using UKF. These results are to some extent comparable with the work of Koller and Lerner in [2000] in which the generic PF algorithm was applied to the watertank problem using the network structure in Figure 2. We have shown that a different network structure improved the accuracy of PF (PFMC) and still the use of UKF significantly improved the ability to track the true failure nodes and estimate the continuous state variables with a low number of samples.

All in all, we have compared several inference techniques and shown that it is possible to do inference in a complex hybrid DBN. All in all, we conclude that we should choose PFUKF, when the measurements are noisy (i.e. when UKF is not able to make reliable estimates), when we do not know the 'true' process and measurement models, when we work on higher order models (data not shown) and have the necessary computational time. Otherwise, we would settle for UKF or standard PF. However, PF has the disadvantage that is network structure sensitive.

5. REFERENCES

[Andersen & Andersen, 2003] M.N. Andersen and R.Ø. Andersen. *Filtering in Hybrid Dynamic Bayesian Networks* Master thesis, IMM, Techinical University of Denmark, 2003. http://www.imm.dtu.dk/ mna/public.php

[Doucet et al., 1998] A. Doucet, S. Godsill and C. Andrieu. On Sequential Monte Carlo Sampling Methods for Bayesian Filtering. Technical report CUED/F-INFENG/TR 310, Dept. of Eng., Cambridge University, 1998.

[Julier & Uhlmann, 1997] S.J. Julier and J.K. Uhlmann. A new extension of the Kalman filter to nonlinear systems. In *Proceedings of AeroSense: The 11'th International Symposiom on Aerospace/Defence Sensing, Simulation and Controls*, Multi Sensor Fusion, Tracking and Resource Management II. Orlando, Florida, 1997.

[Koller & Lerner, 2000] D. Koller and U. Lerner. *Sampling in Factored Dynamic Systems*. In Sequential Monte Carlo Methods in Practice: pages 445-464, Springer-Verlag, New York, 2000.

[v.d. Merwe et al., 2000] R. v.d. Merwe, A. Doucet, N. de Freitas and E. Wan. The Unscented Particle Filter. Technical report CUED/F-INFENG/TR-380, Dept. of Eng., Cambridge University, 2000.